A fast algorithm for high-order sparse linear prediction

Tobias Lindstrøm Jensen[1], Daniele Giacobello[2], Toon van Waterschoot[3], Mads Græsbøll Christensen[4]

[1]Signal and Information Processing, Aalborg Universitet, Denmark

[2]Codec Technologies R&D, DTS Inc., Calabasas, CA, USA

[3]Department of Electrical Engineering (ESAT-STADIUS/ETC), KU Leuven, Belgium

[4]Audio Analysis Lab., AD:MT, Aalborg Universitet, Denmark

tlj@es.aau.dk, giacobello@ieee.org, toon.vanwaterschoot@esat.kuleuven.be, mgc@create.aau.dk

***Abstract*— Using a sparsity promoting convex penalty function on high-order linear prediction coefficients and residuals addresses some inherent limitations of standard linear prediction methods. This formulation, however, is computationally more demanding which may limit its use, in particular for embedded signal processing. We show that the matrix structures associated with an alternating direction method of multipliers algorithm for solving the high-order sparse linear prediction problem are more tractable than the matrix structures for interior-point methods and that a few tens of iterations suffice to achieve similar results, in terms of prediction gain, as an interior-point method.**

1 Background

Sparse linear prediction (SLP) [1, 2] revisits the linear prediction (LP) framework [3, 4] in light of the developments that took place in the recent years in the field of convex optimization and sparse representations. SLP has proved to be an interesting alternative to classic LP by allowing better statistical models and more meaningful signal representation finding its way in various applications, e.g., [5–7]. While software packages like CVX+SeDuMi [8, 9] allow to quickly reproduce the SLP algorithm [10, 11], serious efforts to make SLP and other algorithms requiring convex optimization run faster and, possibly, in a real-time platform, is a current matter of research in signal processing [12, 13].

LP provides a compact representation for the signal $x[t]$ as:

$$x[t] = \sum_{n=1}^{N} \alpha_n x[t-n] + r[t], \qquad (1)$$

where $\alpha = [\alpha_n]_{n=1}^{N}$ are the prediction coefficients and $r[t]$ is the prediction error. A common route for estimation of α is via:

$$\underset{\alpha}{\text{minimize}}\ \|x - X\alpha\|_p^p \qquad (2)$$

where $\|\cdot\|_p$ is the p-norm and we here are working with a vectorized version of (1) over a certain frame. With $p = 2$, a closed-form solution can be obtained as $\alpha = (X^T X)^{-1} X^T x$.

The LP model finds one of its most successful applications in speech and audio processing [4]. However, particularly in speech processing, traditional LP fails to provide a general framework when signal redundancies are present at different time intervals. This is the case where a given segment x has short-term and long-term redundancies and cannot be represented by a simple linear prediction model with a limited number of taps. Traditional approaches tend to represent short-term redundancies using traditional LP and represent long-term redundancies by applying a so-called long-term predictor (LTP) with a very limited number of taps clustered around the *pitch period* of the speech or audio signal. Since the combination of these two filters is a high-order sparse predictor, a more effective way to model these types of signal was shown by increasing the order of the predictor and apply a sparsity criterion on its coefficients [14]. In addition, by applying the 1-norm also on the residual, both modeling and coding advantages can be achieved [2]. A SLP formulation then becomes:

$$\underset{\alpha}{\text{minimize}}\ f(\alpha) = \|x - X\alpha\|_1 + \gamma\|\alpha\|_1 . \qquad (3)$$

Solving (3) is however more complex than traditional 2-norm based LP and state-of-the-art methods for real-time optimization have, to some extent, focused on code generation based on interior-point (IP) methods [12, 15]. The most significant matrix structure these methods exploit is sparsity, i.e., for example at code-generation exploiting $[A\,0][x^T\, y^T]^T = Ax$ such that we avoid computing $0 \cdot y$. However, many applications in signal processing, including the problem (3), are dense. Further, for IP methods the main bulk of work is to solve a linear system of equations in each iteration where additional weighting matrices are introduced and in particular diagonal matrices for linear programming. Such diagonal matrices often do not allow the possibility of faster direct methods for solving a linear system [16]. Specifically, IP methods for the SLP problem (3) would have cubic per iteration time complexity [17]. In particular, the introduction of the diagonal weighting matrix for the SLP problem prohibits the exploration of Toeplitz structure since such matrices do not have a low displacement rank.

Instead of using an IP method we consider an alternating method of multipliers (ADMM) algorithm. The problem in (3) can be recast as the following least absolute deviation problem

$$\underset{\alpha}{\text{minimize}}\ \left\| \begin{bmatrix} \gamma I \\ -X \end{bmatrix} \alpha - \begin{bmatrix} 0 \\ -x \end{bmatrix} \right\|_1 . \qquad (4)$$

for which the ADMM algorithm is [18]:

$$\begin{aligned}
\alpha^{(k+1)} &= (X^T X + \gamma^2 I)^{-1} \begin{bmatrix} \gamma I & -X^T \end{bmatrix} (y^{(k)} + \begin{bmatrix} 0 \\ -x \end{bmatrix} - u^{(k)}) \\
e^{(k+1)} &= x - X\alpha^{(k+1)} \\
y^{(k+1)} &= S_{1/\rho}\left(\begin{bmatrix} \gamma\alpha^{(k+1)} \\ e^{(k+1)} \end{bmatrix} + u^{(k)} \right) \\
u^{(k+1)} &= u^{(k)} + \begin{bmatrix} \gamma\alpha^{(k+1)} \\ e^{(k+1)} \end{bmatrix} - y^{(k+1)} .
\end{aligned}$$

The work of T. L. Jensen is supported by The Danish Council for Independent Research, Technology and Production Sciences, grant number 4005-00122. The work of T. van Waterschoot is supported by the KU Leuven Impulse Fund, grant number IMP/14/037.

Notice that in this ADMM formulation there is no reweighting of X (or X^TX) and we may then use fast (and superfast) methods for solving the linear Toeplitz system [19, 20].

2 Numerical Simulations

For the numerical simulations we will focus on the application of SLP in speech processing, however, we will apply a objective measures and the results is then extendable to other application scenarios. In particular, we investigated the prediction gain performance as a function of the number of iterations of the presented ADMM algorithm and the associated computational cost assessed by timing.

We processed the vowel and semivowel phones [21] from the TIMIT database ($f_s = 16$kHz), belonging to 3696 sentences from 462 speakers. We chose the ones of duration of at least 640 samples (40 ms) for a total of about 40,000 voiced speech frames. In this investigation, we extend the analysis in [22] by investigating the prediction gain from the ADMM solution with a different number of iterations and compare with the IP solution obtained through the CVX+SeDuMi interface and solver. The regularization parameter $\gamma = 0.12$ was obtained through a modified L-curve analysis [23] by using all except 50 frames picked randomly that will be used as a test set. We chose $N = 250$ such that it is possible to cover the pitch lag range $T_p \in [34, 231]$ as done in commercial speech codecs like the wideband version of the Adaptive Multi-Rate codec (AMR-WB [24]). The results for the test set is shown in Figure 1. We can see that at 30 iterations the mean value of the IP solution falls within the 95% confidence interval of the ADMM solution, proving that the two algorithms exhibit statistically the same performance.

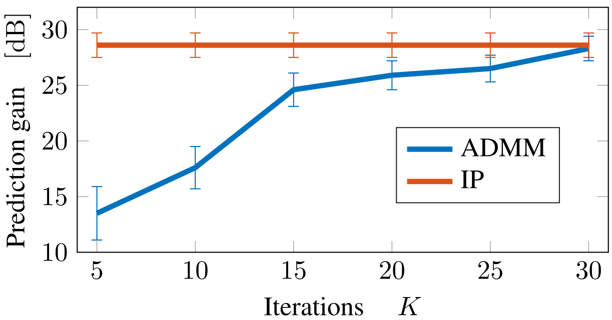

Figure 1: Average prediction gains for a fixed number of iterations for the ADMM solution. A 95% confidence interval is shown. The IP solutions as returned by CVX+SeDuMi is independent of the fixed number of iterations K but shown for the ease of comparison.

Using a C++ implementation, we can run $k = 30$ iterations on a standard laptop in approximately 1.7ms in double precision. This is an indication that an ADMM algorithm for the SLP problem may be viable for real-time and embedded optimization but further algorithm design investigations are necessary to address this possibility. We note that we are applying the ADMM algorithm in its straightforward form but several variants and extensions may be useful for solving the sparse linear prediction problem efficiently and should be considered for further investigations. On particular choice is a preconditioned ADMM where the algorithm does not involve solving a linear system of equation [25].

References

[1] D. Giacobello, M. G. Christensen, J. Dahl, S. H. Jensen, and M. Moonen, "Sparse linear predictors for speech processing," in *Proc. INTERSPEECH*, pp. 1353–1356, 2008.

[2] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen, and M. Moonen, "Sparse linear prediction and its applications to speech processing," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 5, pp. 1644–1657, 2012.

[3] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[4] T. Bäckström, *Linear predictive modeling of speech: constraints and line spectrum pair decomposition*, Helsinki University of Technology, 2004.

[5] D. Angelosante, "Sparse regressions for joint segmentation and linear prediction," in *Proc. ICASSP*, pp. 335–339, 2014.

[6] A. Jukic, T. van Waterschoot, T. Gerkmann, and S. Doclo, "Multi-channel linear prediction-based speech dereverberation with sparse priors," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 9, pp. 1509–1520, 2015.

[7] J. Koloda, J. Østergaard, S.H. Jensen, V. Sanchez, and A.M. Peinado, "Sequential error concealment for video/images by sparse linear prediction," *IEEE Trans. Multimedia*, vol. 15, no. 4, pp. 957–969, 2013.

[8] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, v. 1.21," `http://cvxr.com/cvx/`, 2011.

[9] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Methods Software*, vol. 11-12, pp. 625–653, 1999.

[10] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen, and M. Moonen, "Speech coding based on sparse linear prediction," in *Proc. EUSIPCO*, pp. 2524–2528, 2009.

[11] D. Giacobello, T. van Waterschoot, M. G. Christensen, S. H. Jensen, and M. Moonen, "High-order sparse linear predictors for audio processing," in *Proc. EUSIPCO*, pp. 234–238, 2010.

[12] J. Mattingley and S. Boyd, "Real-time convex optimization in signal processing," *IEEE Signal Process. Magazine*, vol. 27, no. 3, pp. 50–61, 2010.

[13] B. Defraene, T. van Waterschoot, H. J. Ferreau, M. Diehl, and M. Moonen, "Real-time perception-based clipping of audio signals using convex optimization," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 10, pp. 2657–2671, 2012.

[14] D. Giacobello, M. G. Christensen, J. Dahl, S. H. Jensen, and M. Moonen, "Joint estimation of short-term and long-term predictors in speech coders," in *Proc. ICASSP*, pp. 4109–4112, 2009.

[15] A. Domahidi, E. Chu, and S. Boyd, "ECOS: An SOCP solver for embedded systems," in *Proc. EUSIPCO*, pp. 3071–3076, 2013.

[16] D. O'Connor and L. Vandenberghe, "Primal-dual decomposition by operator splitting and applications to image deblurring," *SIAM J. Imag. Sci.*, vol. 7, pp. 1724–1754, 2014.

[17] T. L. Jensen, D. Giacobello, M. G. Christensen, S. H. Jensen, and M. Moonen, "Real-time implementations of sparse linear prediction for speech processing," in *Proc. ICASSP*, pp. 8184–8188, 2013.

[18] T. L. Jensen, D. Giacobello, T. van Waterschoot, and M. G. Christensen, "Computation analysis of a fast algorithm for high-order sparse linear prediction," in *Proc. EUSIPCO*, 2016.

[19] R. R. Bitmead and B. D.O. Anderson, "Asymptotically fast solution of Toeplitz and related systems of linear equations," *Linear Algebra Appl.*, vol. 34, pp. 103–116, 1980.

[20] G. S. Ammar and W. B. Gragg, "Superfast solution of real positive definite Toeplitz systems," *SIAM J. Matrix Analysis. Appl.*, vol. 9, no. 1, pp. 61–76, 1988.

[21] A. K. Halberstadt and J. R. Glass, "Heterogeneous acoustic measurements for phonetic classification," in *EUROSPEECH*, 1997.

[22] T. L. Jensen, D. Giacobello, T. van Waterschoot, and M. G. Christensen, "Fast algorithms for high-order sparse linear prediction with applications to speech processing," *Speech Comm.*, vol. 76, pp. 143–156, 2016.

[23] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.*, vol. 14, no. 6, pp. 1487–1503, 1993.

[24] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jarvinen, "The adaptive multirate wideband speech codec (AMR-WB)," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 8, pp. 620–636, 2002.

[25] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, 2011.